Projet RESEDA: Rapport sur les recherches effectuées du 1er oct. 1975 au 1er avril 1976

(Rap. / C.N.R.S. / E.R.H.F. / 1976 / D.G.R.S.T.-1) *

ABRÉGÉ: 1. *Préliminaires*. – 2. *Organisation et codage des informations*. – 3. *Questions et inférences*. – 4. *Choix du langage de programmation*.

1. PRÉLIMINAIRES

1.1. Le projet RESEDA a été conçu pour fournir à des spécialistes du XIVe et XVe siècle un mode de traitement conversationnel des corpus d'informations biographiques.

L'intérêt méthodologique du projet nous semble résider essentiellement dans la tentative d'enrichir les techniques de documentation automatique par des instruments dotés d'une certaine puissance déductive (Cros *et al.* 1968; Fossier, Créhange 1970), et d'évaluer ainsi sur un exemple concret le poids des outils à mettre en place pour simuler un ensemble de démarches intellectuelles d'une certaine complexité.

D'autre part, le recours à des procédures très sophistiquées et presque imposé par le caractère extrêmement complexe et très souvent implicite des relations interpersonnelles considérées. Le but visé est en fait la création d'un système qui n'ait pas seulement une fonction statique de récupération d'une information stockée mais aussi une fonction dynamique, permettant d'établir des liaisons nouvelles entre les données et d'accroître en quelque sorte les connaissances de départ. RESEDA doit donc s'engager dans une certaine mesure sur le domaine de l'intelligence artificielle.

---

* Le projet en question, réalisé par l'Equipe de Recherche sur l'Humanisme Français des XIVe et XVe siècles (E.R.H.F.) grâce au financement conjoint du Centre National de la Recherche Scientifique (C.N.R.S.) et de la Délégation Générale à la Recherche Scientifique et Technique (D.G.R.S.T.) – convention 75-7-0456 – est placé sous la responsabilité scientifique de M. Gilbert Ouy, maître de recherche, directeur de l'E.R.H.F. L'organisation générale de RESEDA et la planification du travail reviennent à M. Gian Piero Zarri, ingénieur-docteur, assisté, pour l'étude du métalanguage, de Mme Monique Ornato, documentaliste. Pour octobre 1977, est prévue la réalisation d'un premier modéle opérationnel de RESEDA portant sur les milieux parisiens liés à la « Cour Amoureuse » dite de Charles VI; les renseignements biographiques seront fournis par Mlle Carla Bozzolo qui prépare un ouvrage sur ce sujet.

1.2. Il convient d'écarter immédiatement toute équivoque sur notre façon d'envisager l'usage de cette nouvelle discipline. En effet, notre approche est bien différente de celle d'un Schank (Schank 1973; Schank *et al.* 1975) ou d'un Minsky (Minsky 1974) – pour ne citer que deux des noms les plus connus –. Pour le premier, il s'agit de mettre au point une théorie des phénomènes de conceptualisation sous-jacents aux structures de surface du langage naturel afin d'arriver à sa compréhension par des moyens automatiques. Quant à Minsky, son système de « frames » a été créé pour s'adapter à n'importe quelle situation, tant dans sa perception que dans sa formulation linguistique, et il se situe au niveau de généralité qui permet de prendre en compte *tous* les aspects d'une certaine réalité.

RESEDA est tout à fait étranger à cette problématique très générale; il évite soigneusement, contrairement aux systèmes que l'on vient d'évoquer, d'avancer des hypothèses sur le fonctionnement du mécanisme humain de pensée. Aussi, toutes nos références à toute théorie fondée sur une certaine analyse des processus mentaux (cf. par ex. la théorie des corrélateurs de Ceccato, Ceccato 1962), n'implique de notre part qu'un interêt pragmatique pour un instrument utile sans adhésion ni prise de position sur des fondements théoriques. RESEDA, bien situé dans son contexte humaniste, ne se pose aucun problème de Weltanschauung.

Pour arriver au but que nous nous sommes fixé, ce que nous devons emprunter à l'intelligence artificielle ce sont: 1) pour la description des données de base, le modalités du type « case systems » très riches de possibilités déductives et libéreés de toute fidélité formelle à l'expression en langage naturel (Schank 1973, 1975; Bruce 1975); 2) pour l'utilisation de cette richesse descriptive, un certain nombre de procédures de déduction naturelle, appuyées sur des principes tels que l'appartenance à une classe, l'analogie, la transitivité, etc. (Schwarcz *et al.* 1970; Carbonnel, Collins 1974) ou, de façon plus ambitieuse, fondées sur le recensement des propriétés constitutives de certaines situations (Charniak 1972, 1975; Minsky 1974).

1.3. Avant d'entrer dans le détails du travail effectué durant ces six mois, il n'est peut-être pas inutile de donner d'abord une vue d'esemble de RESEDA.

1.3.1. RESEDA est alimenté de façon continue par les informations biographiques tirées de dépouillements effectués sur les sources les plus variées: documents d'archives, revues, ouvrages érudits, etc... Ces éléments, qui constituent les données brutes sur lesquelles le système va ensuite opérer, sont donc dépourvus *a priori* de toute organisation logique susceptible de les mettre en relation entre eux.

Le principe qui est à la base du stockage des informations est de coder séparément chaque « épisode », c'est-à-dire chaque unité logique où se trouve relatée une « action » – ou un « état » – référée à un ou plusieurs

« personnages » et caractérisée par un lieu-temps déterminé. Ainsi l'information:

« de Paris, Clamanges s'était rendu en Avignon. Il y retrouva Jean Muret et Jean Moccia. Le 13 novembre 1397, il y fut nommé secrétaire pontifical par Benoît XIII et le resta jusqu'en mai 1398 »
fera l'objet de trois épisodes codés: le voyage de Clamanges, le séjour contemporain des trois hommes à Avignon, la nomination de Clamanges comme secrétaire.

1.3.2. La nature du codage choisi et sa justification, les details du métalangage seront explicités au chapitre 2; il suffira ici de dire que la transcription de chaque épisode constitue ce que nous avons appelé un « plan » par analogie avec les définitions de dictionnaire de Quillian (Quillian 1968; voir aussi Scragg 1975). Pour le codage des plans, RESEDA ne prévoit pas la mise en place d'un analyseur automatique des sources dépouillées: sa réalisation semble utopique – au moins dans une première étape – vu la nature très érudite de ces sources et une certaine complexité du métalangage. Par contre, il paraît indispensable de réaliser une série de programmes conversationnels qui guideront l'analyste dans le codage. Ce prompting signalera à chaque pas les fautes de syntaxe formelle, les impropriétés de lexique, et d'une manière générale proposera les démarches autorisées.

Les plans sont doublement ordonnés entre eux. L'organisation physique est réalisée selon un critère chronologique *sui generis*, spécialement élaboré pour tenir compte des incertitudes et des imprécisions propres à la documentation médiévale. L'organisation logique distribue les plans entre différents « volumes », chaque volume étant consacré à un seul personnage, la « vedette ». Le critère d'attribution d'un plan donné à un volume est la présence de la vedette dans la formulation de l'épisode que le plan relate. Etant donné que l'épisode peut concerner plusieurs personnages à la fois, c'est le plan entier qui sera attribué aux différents volumes correspondants (Scragg 1975). Chaque volume peut donc être considéré comme une liste d'adresses qui pointent vers les étiquettes des plans (classement inversé). Ainsi le plan du « voyage de Clamanges » sera répertorié dans le seul volume de celui-ci; le plan du « séjour contemporain des trois hommes » dans les volumes de Clamanges, Muret et Moccia; le plan de « la nomination » dans les volumes de Clamanges et du pape Benoît.

On peut évoquer alors le problème, bien caractéristique de la documentation médiévale, des informations controversées: chacune des versions contradictoires doit faire l'objet d'un plan indépendant, mais la série des plans concernant la controverse est reprise dans une liste spéciale. Quand une information obtenue à la suite d'une question posée à RESEDA est comprise dans un plan ainsi listé, on doit répéter la même stratégie de recerche qui a permis de sélectionner ce plan en l'appliquant à chacun de ceux qui lui sont associés.

1.3.3. Les stratégies de recherche employées par RESEDA vont de la simple récupération des informations stockées à des formes plus complexes demandant des opérations de généralisation ou l'exploration de structures *sui generis*, etc. (cf. 3.2.2.). Mais la démarche peut-être la plus intéressante est celle qui devrait permettre d'établir automatiquement des liens, de type surtout causal, entre les épisodes (« the most important conceptual relation is that of causality », Schank 1973). Initialement, ceux-ci apparaissent en effet comme un nuage de points disjoints que ne vient organiser aucun réseau de relations. L'existence de liens préalablement connus entre certains épisodes – et qui ont évidemment été enregistrés dès l'entrée – ne change rien à la généralité du problème.

Le réseau que l'on va donc développer au fur et à mesure de l'utilisation de RESEDA sera mémorisé de façon permanente afin de mettre à la disposition de l'utilisateur une quantité d'information grandissante tout en diminuant l'entropie du système. Dans cet esprit, voici un exemple suffisamment complexe de ce que l'on a l'intention de demander à RESEDA; nous n'en donnons ici que les grandes lignes car il sera repris en détail (cf. 3.2.3.). Supposons une question du type a) « Quelle est l'attitude de Pierre d'Ailly envers l'Université de Paris au cours de l'année 1395? ». Dans le déroulement de la recherche ainsi déclenchée, à défaut d'une réponse directe, on sera amené à retenir comme étape intermédiaire l'information b) « le 2 août 1395, l'Université de Paris interdit la présence de Pierre d'Ailly à ses assemblées sur le schisme », où les rôles respectifs des deux vedettes dans la question sont intervertis (dans a) et dans b), les « actions » relèvent du même « concept »: manifestation physique ou intellectuelle à l'égard de quelqu'un). b) fournit une indication supplémentaire au sujet d'un « argoment » possible de controverse: le schisme. On recherche alors une manifestation antérieure de Pierre d'Ailly sur ce sujet sans exiger cette fois l'Université de Paris come interlocuteur: on trouve c) « vers avril 1395, Pierre d'Ailly prend ouvertement position en faveur du pape d'Avignon Benoît XIII au sujet du schisme ». Il reste à vérifier si le pape Benoît a quelque rapport antagoniste avec l'Université au sujet du schisme et l'on trouve d) « dès le 6 janvier 1391 et jusqu'en 1418, l'Université de Paris s'oppose au pape d'Avignon au sujet du schisme ».

A elles trois, ces informations fournissent la réponse; l'utilisateur a tous les éléments pour déduire que, vers avril 1395, Pierre d'Ailly prend une position contraire à celle de l'Université au sujet du schisme. Selon l'hypothèse qu'une « manifestation » de A « à l'égard » de B peut être causée par une « manifestation » antérieure de B sur le même sujet, on sera amené à retenir qui c) – où Pierre d'Ailly, objet de l'interdiction dans b), figure cette fois comme sujet – est la « cause » de b); un lien de causalité sera donc instauré de c) vers b).

1.3.4. Deux remarques s'imposent pour conclure. Tout d'abord, la démarche qui vient d'être exposée pourra donner l'impression d'être imaginée

pour les besoins de la cause. Et, en effet, la restriction de domaine effectuée par RESEDA – et qui a déjà été amplement commentée – fait que l'on aura toujours des procédures strictement liées à un sous-ensemble délimité du contexte socio-historique médiéval.

Dans les termes de Minsky, nous ne mettons en oeuvre qu'un seul « frame-system », celui de la « situation politico-religieuse de la France à la fin de la guerre de 100 ans ». Toutefois, le niveau de précision attendu des réponses de RESEDA implique une utilisation très raffinée des ressources informationnelles de cet unique frame-system et apport donc à ses procédures un intérêt qui devrait déborder le domaine choisi.

Enfin, la réponse de l'exemple précédent est le produit caractérisé des structures d'inférences que – par opposition à celles plus élémentaires dont il a été question auparavant – on peut appeler de 2e degré. Il s'agit en effet de résultats d'un type « probable » plutôt que « sûr », où le caractère d'incertitude est lié aux inconvénients de notre vision incomplète du monde médiéval et bien entendu aux difficultés générales inhérentes à la mise en place de tout raisonnement mécanique « intelligent ». De telles réponses devront donc toujours être soumises à une mesure de validation avant d'être définitivement acceptées pour donner lieu à l'enregistrement d'une nouvelle information. Cette validation ne saurait être mécanisée que pour la partie des contrôles les plus routiniers, la partie essentielle reposant plutôt sur la compétence historique du chercheur.

C'est pourquoi, sur le plan des principes au moins, on sera amené à maintenir la distinction traditionnelle entre « questions d'utilisateur » proprement dites qui ne donnent accès qu'aux informations déjà mémorisées et « questions de système » réservées à l'enrichissement de RESEDA.

2. Organisation et codage des informations

2.1. RESEDA s'organise autour d'une structure de mémoire à deux niveaux: les informations « personnages », qui constituent la base de données au sens habituel du terme, continuellement remise à jour au cours du processus d'alimentation du réseau, et le « métalangage », choisi de façon pragmatique en fonction de nos objectifs et qui sert de support à l'expression des informations.

2.1.1. Les informations « personnages » font l'objet des « plans » et des « volumes » qui ont été présentés au chapitre précédent. Elles sont donc introduites en mémoire au moment du codage en plan des épisodes.

2.1.1.1. Pour chaque plan, ce codage est divisé en deux parties, le cadre circonstanciel ou « thématique » et la description proprement dite. Ce

découpage n'a rien d'arbitraire: la nature « historique » de nos informations rend inévitables les références de temps et de lieu qu'il sera donc toujours possible d'isoler. Du point de vue plus strictement pratique, cette forme de présentation permet d'alléger le codage en évitant l'introduction de « corrélateurs » (voir plus bas) particuliers pour la localisation et la datation. De plus, on trouvera tout de suite, sans entrer dans les détails de l'épisode, ces deux types d'informations qui, vu le domaine de RESEDA, sont très souvent utilisés pour diriger la recherche.

2.1.1.1.1. La thématique, avec 7 zones d'enregistrement, séparées dans les exemples qui suivent par des barres obliques, prend donc en charge les indications de datation et localisation de l'épisode et aussi l'identification de l'informateur (le plus souvent des références bibliographiques). Les trois premières zones permettent de distinguer en: date de début, dates d'attestation, date de fin; les trois zones suivantes en: lieu de départ, lieux de passage, lieu d'arrivée ou lieu tout court. Par date « d'attestation », on entend le constat qu'à ce moment donné l'action – ou l'état – dont il est fait mention dans l'épisode est en cours, selon le témoignage de la source dépouillée (par exemple un document de chancellerie pontificale faisant état d'une requête de Clamanges le 13 décembre 1394, signale en même temps qu'il était alors chapelain de Notre Dame de Châlons). En ce qui concerne le codage de la localisation, l'organisation en trois zones a évidemment été adoptée de façon à pouvoir prendre en compte les déplacements. Lorsque l'épisode se déroule dans un lieu unique (cf. Notre Dame de Châlons dans l'exemple précedent) les quatrième et cinquième zones restent vides tandis que la sixième contient le « lieu tout court ».

Sans entrer davantage dans le détail des combinaisons possibles, soulignons simplement qu'aucune de ces six premières zones n'est obligatoirement remplie, contrairement à la septième et dernière qui, elle, doit toujours donner l'origine de l'information relatée dans l'épisode. Les indications de datation de la thématique servent pour l'organisation chronologique des plans dont il a déjà été question.

2.1.1.1.2. La description de l'épisode – résolument indépendante de sa formulation en langage naturel – est réalisée par une représentation arboriforme, dont les noeuds sont des « corrélateurs » (Ceccato 1962) et les terminaux des « prédicats » – éventuellement précédés de « modulateurs » –, des éléments du « lexique » ou des vedettes (les « corrélateurs », « prédicats », « modulateurs » et termes du « lexique » sont des éléments du métalangage). Ce mode de représentation rejoint certains résultats des linguistes qui travaillent sur la compréhension du langage naturel avec des outils du type « conceptual case systems ».

2.1.1.1.2.1. Pour illustrer l'interaction dans le codage de ces différents éléments on peut reprendre de façon formalisée le paragraphe de biographie

de Clamanges donné en exemple au chapitre précédent (le codage n'est que partiel pour d'évidentes raisons de clarté):

1) thèmatique: /-/-/avant-13-novembre-1397/Paris/-/Avignon/
Bibl: Glorieux/
description: (*CLAMANGES* .SUJ. (**DEPLACER** .OBJ. *CLAMANGES*))
2) /-/13- novembre-1397/-/-/-/Avignon/Bibl.: Glorieux/
(((*CLAMANGES* .COORD. *MURET*) .COORD. *MOCCIA*) .SUJ. ***E/RESIDANT***)
3) /13-novembre-1397/-/mai-1398/-/-/Avignon/Bibl: Glorieux/
(*CLAMANGES* .SUJ. ((***soc* + *E/BENEFICIANT*** .SPECIF. *secrétariat-pontifical*) .SOURCE. *BENOIT-XIII*))

où

— Glorieux, nom de l'auteur de la biographie, symbolise ici un codage qui prend en compte la nature de l'ouvrage, son auteur, son titre, etc.
— SUJ, OBJ, COORD, SOURCE sont des « corrélateurs ».
— ***DEPLACER*** est un « prédicat d'action », ***E/BENEFICIANT***, ***E/RESIDANT*** des « prédicats d'états » et ***soc*** un « modulateur ».
— *CLAMANGES, MURET, MOCCIA, BENOIT-XIII* sont des vedettes.
— *secrétariat-pontifical* est un terme du « lexique ».

2.1.1.1.2.2. Il est bien évident que les termes du vocabulaire français employés ci-dessus n'ont qu'une valeur mnémonique pour guider l'analiste dans la rédaction des épisodes, leur « signification » est purement interne au système (voir plus bas). Par ailleurs, insistons encore sur le caractère purement pragmatique de la classification du métalangage: la définition des différentes catégories n'est rien d'autre qu'une définition par liste.

Par exemple, seront « corrélateurs » toutes les unités comprises dans la série close: « ARGUMENT, CAUSE, CONDITION, COORDINATION, DESTINATION, FINALITE, MODALITE, OBJET, QUALIFICATION, RELATION, SOURCE, SPECIFICATION, SUJET ».

En ce qui concerne la fonction logique de ces différentes catégories, elle est suffisamment claire en soi pour les corrélateurs – qui nous permettent d'organiser systématiquement le codage en groupes de trois éléments emboitables –, pour les prédicats – correspondant en gros aux verbes du langage naturel –, pour le lexique et les vedettes. Les modulateurs ont pour rôle de restreindre au maximum la liste des prédicats en permettant de préciser le domaine de l'action ou de l'état relaté dans l'épisode per le biais d'un nombre limité de préfixes (voir aussi Mel'chuk, Zholkovskij 1970). ***E/BENEFICIANT*** a le sens général « jouir de quelque chose pendant une certaine période »; avec la préfixe ***soc*** (social) comme dans l'exemple précédant, cette jouissance vient s'inscrire dans le cadre des inter-relations institutionnelles de l'époque; avec le préfixe ***ment*** (mental) par contre, c'est un enrichissement intellectuel qui est manifesté.

2.1.1.2. Un système de « pointeurs » est associé aux plans pour mémoriser dans RESEDA des informations sur les relations qui existent entre des épisodes relatés dans des plans différents. Ces pointeurs seront introduits soit au moment du codage et de la mise à jour des données soit comme résultat des inférences éventuellement déclenchées par certaines questions de système (cf. chapitre précédent). On établit ainsi des « raccourcis » qui permettent d'associer directement entre elles certaines informations pour lesquelles la première mise en relation a nécessité la création d'une longue chaîne d'inférences (Quillian 1968).

Les pointeurs, du point de vue conceptuel, ne sont en fait que des corrélateurs particuliers. Dans l'organisation actuelle de RESEDA, les corrélateurs CAUSE et FINAL sont réservés à cette unique fonction de mise en relation d'épisodes séparés; COORD, qui apparaît normalement à l'intérieur des plans codés, a dans ce cas un rôle auxiliaire.

Il faut s'expliquer sur la coexistence de FINAL et de CAUSE dans une fonction apparemment symétrique. En fait, ces deux corrélateurs se distinguent d'une façon toute pragmatique qui permet de réserver FINAL pour introduire certains épisodes qui – dans l'état des connaissances au moment du codage – apparaissent comme les visées, sans réalisation confirmée, du prédicat exprimé dans le premier corrélé. Les rôles de ces deux corrélateurs trouvent une illustration dans cette suite d'épisodes:

4) /après-11-août-1408/-/-/Paris/-/Cambrai/Bibl: Valois/
(*CHARLES-VI* .SUJ. (**DEPLACER** .OBJ. (*SAINT-POL* .QUALIF. *comte*)))
5) /après-11-août-1408/-/après-11-août-1408/-/-/Cambrai/Bibl: Valois/
(*SAINT-POL* .SUJ. (**contre + exec + PRENDRE-POSITION** .OBJ. *PIERRE-D'AILLY*))
6) /après-11-août-1408/-/-/-/-/Cambrai/Bibl: Valois/
(*PIERRE D'AILLY* .SUJ. **neg + phys + E/LIBRE**)
7) /-/11-août-1408/-/-/-/Paris/Bibl: Valois/
(*PIERRE D'AILLY* .SUJ. (**neg + E/PRESENT** .SPECIF. *5e-concile-Eglise-de-France*))
((4 .FINAL. (5 .FINAL. 6)) .CAUSE. 7)

dont la signification en langage naturel est la suivante:

« constatant l'absence de Pierre d'Ailly, le 11 août 1408, à l'ouverture du 5e concile de l'Eglise de France à Paris, Charles VI envoya le comte de St Pol l'arrêter à Cambrai ».

Notons que le modulateur *exec* indique que l'action est « exécutée » sur l'ordre d'un responsabile autre que le sujet.

Du fait du caractère « non-confirmé » donné à l'épisode 2e corrélé de FINAL, ce pointeur ne peut en aucun cas être le résultat d'une démarche d'inférence, ce qui dans RESEDA est réservé à la mise en relation d'épi-

sodes dont la réalité historique est déjà établie. Par contre, si un élément nouveau apporte un témoignage de la réalisation effective de l'épisode 2e corrélé, il faudra, du fait du principe qui vient d'être énoncé, remplacer FINAL par CAUSE en inversant l'ordre des corrélés.

2.1.2. Les constituants du métalangage, qui viennent d'être présentés dans leur fonctionnement, peuvent être classés en deux catégories par rapport à leur représentation informatique.

2.1.2.1. Dans la premiére, on range ceux dont la fonction sémantique s'exprime sous forme de programmes (Winograd 1972, 1973). Les « corrélateurs », les « prédicats » et les « modulateurs » entrent dans cette catégorie. Chaque prédicat a comme référent un programme particulier qui en décrit le mode d'emploi en termes de procédure; modulateurs et corrélateurs y jouent le rôle d'éléments différenciateurs en aiguillant la recherche vers des subroutines différentes.

Par exemple, dans le cadre général d'un « mouvement d'une position de départ à une position d'arrivée », propre au programme ***DEPLACER***, la présence éventuelle d'un modulateur comme ***ment*** permet d'écarter les embranchements correspondant aux significations hors de cause, comme « voyager », « changer de poste », etc. Le programme s'attend maintenant à se trouver dans le schéma de base suivant:

(*V1* .SUJ. ((***ment*** + ***DEPLACER*** .ARG. *a*) .OBJ. *V2*));

a définit le thème du changement d'attitude mentale. *V2* – qui dans ce schéma est toujours une vedette comme *V1* – fournit le protagoniste du changement et le programme continue en cherchant dans le volume de ce dernier deux plans chronologiquement ordonnés qui attestent ce mouvement (prédicat ***PRENDRE-POSITION*** diversement modulé) et que l'on peut donc faire coïncider avec la position de départ et celle d'arrivée.

Si le résultat est positif, un nouvel aiguillage sera déterminé par le résultat d'une confrontation entre *V1* et *V2*. En cas de coïncidence, le sous-programme mis en marche correspond au sens « changer d'avis », et sinon au sens « influencer ». Ce dernier prévoit alors la recherche dans le volume de *V1* d'un plan portant une date antérieure à celle du plan « position d'arrivée » et où l'on reconnait l'équivalence de la partie prédicative des épisodes en utilisant, le cas échéant, les définitions de termes de lexiques sous la responsabilité du programme ***PRENDRE-POSITION*** (cf. 3.2.3.).

En cas de succès, un dernier contrôle vérifie la vraisemblance d'une « influence » de *V1* sur *V2*. Actuellement, cette vérification exige simplement l'existence d'au moins un épisode, commun aux volumes de *V1* et *V2*, chronologiquement antérieur au plan « position d'arrivée ». On espère raffiner ultérieurement la notion de contact entre les deux vedettes si l'on arrive à prendre en compte des informations significatives sur le milieu de travail, la résidence, la parenté etc.

Tous les éléments que le programme ***DEPLACER*** permet de recueillir pendant son déroulement – y compris les indications éventuelles d'échec activées aux différents niveaux de la recherche – seront mis à la disposition de l'algorithme général de traitement des questions qui utilisera ces renseignements en fonction du type de question posée.

2.1.2.2. Dans la deuxième catégorie on range les éléments du lexique qui au contraire, exigent d'être définis les uns par rapport aux autres. Cette mise en relation peut prendre trois aspects.

2.1.2.2.1. On a d'abord les structures arboriformes comme dans la traditionnelle relation du général au particulier (ex. classification ecclésiastique des localités en paroisse, doyenné, archidiaconé, diocèse, province). Il est possible d'assurer le passage d'une arborescence à l'autre en instaurant entre certains de leurs noeuds des liens d'« équivalence ». On peut remarquer qu'un certain nombre de ces termes correspondent à un « collectif de vedettes » caractérisé par l'énumération des individus qui le composent. Par exemple, le générique *papes-de-Rome* a pour spécifiques *URBAIN-VI, BONIFACE-IX, INNOCENT-VII, GREGOIRE-XII, MARTIN-V*. Dans ce seul cas, on insère dans le volume de chacune des vedettes concernées tous les plans qui contiennent le terme « collectif ».

2.1.2.2.2. On a ensuite les définitions du type « rubrique de dictionnaire », où sont utilisés un ou plusieurs plans reliés, au besoin, par des corrélateurs-pointeurs. On peut donner en exemple le noyau central de la définition de *grand-schisme-d'occident*:

8) /1378/-/1418/-/-/occident/Bibl: consensus/
(*église-catholique* .SUJ. ((***contre*** + ***PRENDRE-POSITION*** .OBJ. *église-catholique*) .ARG. *papauté*))
9) /-/9-avril-1378/-/-/-/Rome/Bibl: consensus/
(*sacré-collège-1* .SUJ. ((***soc*** + ***DONNER*** .OBJ. *papauté*) .DEST. *URBAIN-VI*))
10) /-/21-septembre-1378/-/-/-/Fondi/Bibl: consensus/
(*sacré-collège-2* .SUJ. ((***soc*** + ***DONNER*** .OBJ. *papauté*) .DEST. *CLEMENT-VII*)).
(8 .CAUSE. (9 .COORD. 10))

La signification en langage naturel de ce noyau de définition est la suivante:

« de 1378 à 1418, en occident, l'Eglise est divisée en elle-même sur l'occupation du siège pontifical. En effet, l'élection à Rome, le 9 avril 1378, du pape Urbain VI par les cardinaux présents du Sacré Collège fut remise en question au cours d'une nouvelle élection réalisée dans la ville voisine de Fondi au bénéfice de Clément VII, le 21 septembre, par un Sacré Collège légèrement modifié ».

Evidemment, pour ces définitions de dictionnaire, les pointeurs ne sont pas le résultat d'inférence mais sont établis d'avance au moment de la constitution de RESEDA.

Lorsqu'il y a plusieurs plans dans une définition, ils ne sont pas nécessairement tous reliés entre eux par des pointeurs. Par exemple, pour *schisme*, on aura, en plus du « noyau », toute une série de plans qui apportent des informations de caractère assez général sur cet évènement et qui font partie soit d'autres définitions soit du volume de certaines vedettes. Dans ce cas, ils contiennent très souvent *schisme* come deuxième corrélé de SPECIF ou de ARG, ce qui permet d'introduire dans la définition des éléments de récursivité. Voici par exemple un plan – déjà cité en clair en 1.3.3. – qui est répertorié aussi bien sous *UNIVERSITE-DE-PARIS* que sous *schisme*:

11) /6-janvier-1391/-/1418/-/-/-/Bibl: consensus/
(*UNIVERSITE-DE-PARIS* .SUJ. ((*contre* + ***PRENDRE-POSITION***
.OBJ. *papes-d'Avignon*) .ARG. *schisme*)).

Dans les cas les plus complexes, les définitions de dictionnaire sont donc représentées, tous comme les vedettes, par des listes d'adresses de plans. La différence fondamentale réside en fait dans la façon de constituer ces « volumes ». Pour les personnages, ils sont constitués automatiquement en empilant, au fur et à mesure de l'alimentation de RESEDA, tous les plans où le personnage est cité. Pour les définitions de dictionnaire, les plans sont délibérément choisis – au moment de l'établissement du système – sur la base de critères d'unité sémantique et de cohérence fonctionnelle; en principe ils ne sont pas susceptibles d'êtres modifiés.

2.1.2.2.3. On a enfin des cas de structuration *sui generis* des éléments du lexique dont l'exemple le plus complexe est fourni par le système de parenté, où différents axes (sexe, mariage, fraternité, filiation) vont se croiser pour déterminer les couples de termes symétriques (du type père-fils, oncle-neveu, belle mère-gendre) et leurs séquences (le fils de la soeur = le neveu). Dans RESEDA, contrairement à la démarche de Lindsay (Lindsay 1973) par exemple, le système de relations de parenté est donné une fois pour toutes au lieu d'être construit progressivement et automatiquement à partir d'un ensemble restreint de données qu'on élargit ensuite au fur et à mesure du développement de la recherche. Cette approche est sans doute intellectuellement plus stimulante mais il faut considérer que l'essentiel pour nous est d'arriver à prendre en compte concrètement les données de parenté sur environ 8 générations plutôt que d'étudier théoriquement la possibilité d'élargir par système les bornes d'une généalogie.

Le corrélateur RELATION ets spécifiquement réservé aux structures *sui generis,* parmi lesquelles ou prévoit actuellement, outre la *parenté*, le *parrainage-stricto-sensu*, le *parrainage-universitaire*, la *protection*.

3. Questions et inférences

3.1. L'utilisateur pour interroger RESEDA formule ses questions de manière indirecte (Carbonnel, Collins 1974) en observant des principes de codage strictement identiques à ceux de l'introduction des plans en mémoire. Pour la rédaction de ces questions, l'utilisateur disposera d'un programme de prompting tout à fait semblable à celui qui guide l'analyste dans le codage des plans (voir 1.3.2.).

Pour spécifier le point d'impact précis de la question, RESEDA offre cinq modalités d'interrogation.

3.1.1. On introduit la marque d'interrogation devant le terme inconnu X qui remplace un corrélé à l'intérieur de la parenthétisation. Si bien que la restitution d'un terme du lexique à l'intérieur de la même structure suffit, dans les cas les plus simples, à transformer la question en une assertion valable.

3.1.2. On introduit la marque d'interrogation devant le terme inconnu X qui remplace le modulateur du prédicat. Il s'agit là d'un usage qui sera illustré en 3.2.3.

3.1.3. On introduit la marque d'interrogation dans une corrélation devant un terme connu du lexique de RESEDA. Cette possibilité correspond en particulier aux recherches de spécifiques dans les arborescences du terme marqué.

3.1.4. On introduit la marque d'interrogation devant l'ensemble de la parenthétisation pour indiquer que la question porte sur la véracité de la formulation avancée.

3.1.5. La marque d'interrogation peut évidemment apparaître aussi dans la thématique. On peut remarquer que RESEDA ne permet pas d'interroger en même temps sur la description et sur le cadre circonstanciel ni – d'une manière plus générale – de poser plus d'une question à la fois.

3.2. RESEDA apporte des réponses de fiabilité variable en utilisant sa structure de mémoire selon différentes stratégies.

3.2.1. La première consiste à rechercher une coïncidence terme à terme entre l'énoncé de la question et tout ou partie de la formulation d'un plan (« direct match »). Cette démarche peut porter tant sur la thématique que sur la description; dans ce dernier cas, l'analogie doit d'abord s'affirmer au niveau des corrélateurs et de leur organisation parenthétique. La seule différence admise sera l'absence dans la question de certaines corrélations contenues dans le plan.

Dans ces conditions, on peut considérer la réponse, accompagnée de ses éventuelles « interprétations divergentes » (cf. le problème des « informations controversées », 1.3.2.), comme « sûre dans l'état actuel des connaissances ».

Au delà du direct match strict, la recherche va se faire selon les possibilités offertes par les différentes organisations (voir 2.1.2.2.) des constituants du métalangage dans RESEDA: d'abord au niveau d'une simple extension du direct match, puis à un niveau d'inférence plus élaboré. Ces possibilités se conjuguent évidemment pour une recherche complexe mais nous les isolerons pour plus de netteté.

3.2.2. Extension du direct match: on peut (en adoptant la terminologie de Schwarcz *et al.* 1970) appeler cette façon de procéder « direct match généralisé ».

3.2.2.1. Un premier type est mis en marche à partir des éléments de la question qui dans le lexique de RESEDA sont rattachés aux structures 2.1.2.2.1.

On peut considérer par exemple la question:

/-/-/avant-1394/-/-/Paris/-/
(*CLAMANGES* .SUJ. (**ment + E/BENEFICIANT** .SPECIF. *?cours-fac-arts*)).

qui signifie: Quels cours Clamanges a-t-il suivi à la faculté des arts de Paris avant 1394? Voici un fragment de l'arborescence *enseignement-public*:

. *enseignement-public*
. . *cours-d'enseignement*
. . . *cours-préparatoire*
. . . . *cours-petites-écoles*
.
.
.
.
. . . . *cours-fac-arts*
. . . . . *morale*
. . . . . *logique*
. . . . . *grammaire*
.
.
.
.
. . . *cours-fac-supérieur*
. . . . *cours-fac-théologie*
.
.
.
.
. . *corporation-de-l'enseignement*
etc.

Au niveau de *cours-d'enseignement* se trouve une relation d'équivalence avec *formation-privée* qui est le sommet d'une autre arborescence:

. *formation-privée*
.. *formation-commerciale*
... *comptabilité*
.
.
.
.. *formation-érudite*
... *grec*
... *hébreu*
etc.

L'exploration des spécifiques du terme *cours-fac-arts*, marqué dans la question, fournit une série d'éléments sur lesquels le direct match doit être tenté et qui constitueront, le cas échéant, des réponses « sûres dans l'état actuel des connaissances » comme par exemple l'indication que Clamanges a suivi un cours de morale.

En cas d'échec et avant d'entamer une autre procédure, on peut explorer, dans les limites de seuil convenables, les noeuds situés dans l'arborescence au dessus du terme marqué et leurs spécifiques. On retiendra par exemple que Clamanges avait suivi des cours de théologie (*cours-fac-théologie*). On pourra également visiter, dans d'autres arborescences, les noeuds reliés au terme de départ ou à ses génériques par une relation d'« équivalence » (Cros *et al.* 1968; Schwarcz *et al.* 1970). On pourra trouver alors que Clamanges avait des connaissances en grec (*formation-privée*, *grec*). Avec *cours-fac-théologie* et *grec*, ou aura récupéré des indications précieuses sur les centres d'intérêt de Clamanges dans la période envisagée.

Il ne s'agit pas dans ces derniers cas de véritable réponse mais d'un apport d'éléments d'information reliés au thème de la question, ce qui devrait permettre d'aiguiller l'utilisateur vers d'autres questions pertinentes.

3.2.2.2. Un deuxième type est mis en marche à partir des éléments de la question qui dans le lexique de RESEDA sont rattachés à une structure 2.1.2.2.3.

La découverte d'un lien de parenté indirectement fourni revêt une importance toute particulière étant donné le caractère de la documentation (données biographiques) que RESEDA doit traiter.

Ainsi à la recherche du lien de parenté existant entre X et Y, on ne trouve aucune indication X .REL. (parenté) Y mais on trouve dans le volume de X les attestations X .REL. (parenté) A, B, C et dans le volume de Y les attestations Y .REL. (parenté) M, N, O. Partant en même temps des sommets X et Y, on construit donc deux arborescences où les noeuds du pre-

mier niveau sont représentés par A, B, C et M, N, O et où les niveaux successifs sont obtenus de la même façon à partir de chacun des noeuds de premier niveau. Si dans la construction d'un niveau on trouve un noeud appartenant déjà au même arbre, on ne le prend pas en considération, si au contraire il appartient déjà à l'autre arbre, on a trouvé là le point de coïncidence cherché. Un seuil empirique sera établi pour éviter une trop grande extension des arbres. Si Z est le point commun on aura finalement une chaîne comme:

X .mari-de. C .fille-de. C1 . . . .REL. Z .REL. . . . M1 .REL. M .REL. Y

On se sert alors de la structure *sui generis* de la parenté pour traduire successivement chaque couple adjacent de la chaîne dans les termes de parenté correspondants (X mari de la fille de C1=X beau fils de C1 etc.). On retrouvera finalement le lien cherché.

3.2.3. RESEDA prévoit enfin un certain nombre de structures d'inférence « du 2e degré » (cf. 1.3.4.). Ces procédures sont essentiellement liées à la définition des « prédicats » sous forme de programme (cf. 2.1.2.1.). Nous reprenons ici l'exemple proposé en 1.3.3.

La question a) sera codée ainsi:

/début-1395/-/fin-1395/-/-/-/-/
(*PIERRE-D'AILLY* .SUJ. (*?X* + ***PRENDRE-POSITION*** .OBJ. *UNIVERSITE-DE-PARIS*)).

Comme première procédure, RESEDA s'efforce de réaliser un « direct match »: on examinera le volume du sujet dans les limites de temps définis par la thématique, en essayant de trouver un plan ayant la même formulation minimale que la question et un ou plusieurs modulateurs à la place de ***X***. On notera que, si un utilisateur a déjà une idée très précise du contexte de ***PRENDRE-POSITION*** qu'il veut obtenir, il peut faire suivre le ***X*** par une liste des modulateurs à écarter, précédés chacun du signe « moins » (par exemple ***?X — exec + PRENDRE-POSITION***).

Le « direct match » n'ayant pas donné de résultat, on devrait alors essayer un « direct match généralisé » – c'est-à-dire s'efforcer d'utiliser les définitions « arboriformes » du lexique 2.1.2.2.1. – mais les moyens ne nous en sont pas donnés ici (aucun terme de lexique dans la question).

L'algorithme de réponse appelle alors le programme ***PRENDRE-POSITION*** qui vérifie la présence dans la question d'un modulateur (facultatif), de l'« objet » (obligatoire) et de l'« argument » (facultatif). L'absence d'argument et de modulateur nous aiguille vers une subroutine qui demande, en premier lieu, si l'objet est une vedette. C'est le cas ici, on recherche alors les plans qui sont à l'intersection des volumes du sujet et de l'objet – dans la période définie par la thématique – et qui contiennent entre ces deux

vedettes le prédicat ***PRENDRE-POSITION.*** On trouvera dans notre exemple l'information b) sous la forme:

12) /-/2-août-1395/-/-/-/Paris/Bibl: Valois/
(*UNIVERSITE-DE-PARIS* .SUJ. (***contre*** + ***real*** + ***PRENDRE-POSITION*** .OBJ. *PIERRE-D'AILLY*)).

13) /2-août-1395/-/-/-/Paris/Bibl: Valois/
(*PIERRE-D'AILLY* .SUJ. (***neg*** + ***E/PRESENT*** .SPECIF. (*assemblée-universitaire* .ARG. *schisme*)))

(12 .FINAL. 13)

Une recherche est alors lancée par le programme pour retrouver la première occurrence du corrélateur ARG; elle se poursuit, le cas échéant, dans tous les plans reliés au premier par des corrélateurs-pointeurs. Si cette recherche n'aboutissait pas, on sortirait du programme ***PRENDRE-POSITION*** en fournissant l'ensemble des plans trouvés à titre d'éléments d'information. Le programme reconnait alors dans les éléments tirés de 12+13, la première partie d'un schéma:

α) (*V2* .SUJ. ((***pour/contre*** + ***PRENDRE-POSITION*** .OBJ. *V1*) .ARG. *a*))
β) (*V1* .SUJ. ((***pour/contre*** + ***PRENDRE-POSITION*** .OBJ. *V3*) .ARG. *a*))

(α .CAUSE. β)

(voir aussi les « démons » de Charniak, Charniak 1972); le programme essaie alors de le compléter, quitte à revenir en arrière, en cas d'éches, pour passer par la définition du terme de lexique α (pour la justification du lien de causalité entre α et β cf. 1.3.3.). ***PRENDRE-POSITION*** va donc rechercher dans le volume de *V1* (Pierre d'Ailly) un plan portant une date antérieure à celle de α (empruntée à 12+13) et ayant la forme de β. Si cette recherche aboutit, il reste à chercher dans les plans communs à *V3* et à *V2* (Université de Paris) – par une recherche « direct match généralisé – le schéma suivant:

γ) (*V2/V3* .SUJ. ((***pour/contre*** + ***PRENDRE-POSITION*** .OBJ. *V3/V2*) .ARG. *a*).

où la date est également antérieure à celle de α et où le modulateur doit correspondre aux exigences du tableau:

α	β	γ
contre	pour	contre
contre	contre	pour
pour	pour	pour
pour	contre	contre

Ainsi, β sera satisfait par l'information c) sous la forme:

14) /-/vers-avril-1395/-/-/-/Paris/Bibl: Valois/
(*PIERRE-D'AILLY* .SUJ. ((***pour*** + ***PRENDRE-POSITION*** .OBJ. *BENOIT-XIII* .ARG. *schisme*)).

L'intersection des volumes de *V3* et *V2* donne entre autres l'information d) sous la forme:

11) /6-Janvier-1391/-/1418/-/-/-/Bibl: consensus/
(*UNIVERSITE-DE-PARIS* .SUJ. ((***contre*** + ***PRENDRE-POSITION***
.OBJ. *papes-d'Avignon*) .ARG. *schisme*)).

où la procédure de direct match généralisé permet de reconnaître en 11) le schéma γ (voir « collectif de vedettes » en 2.1.2.2.1.).

Toutes les conditions étant satisfaites, le programme ***PRENDRE-POSITION*** se termine et fournit à l'algorithme de traitement des questions tous les plans situés sur le cheminement qui a abouti; ils constitueront donc la réponse à la question posée. Dans la version définitive de RESEDA, on sortira les plans au fur et à mesure de leur découverte en demandant à l'utilisateur de spécifier s'il veut que le système aille jusqu'au bout de l'embranchement choisi ou s'il préfère l'arrêter, soit que l'information obtenue soit suffisante telle quelle, soit que les perspectives s'en révèlent inintéressantes: l'utilisateur doit donc reformuler la question avec plus de précision.

4. Choix du langage de programmation

4.1. Nous nous permettrons, pour conclure, une courte discussion de certains problèmes de réalisation informatique propres à tous les projets qui se réclament en quelque manière de l'intelligence artificielle.

Au cours de ces dernières années, le développement des recherches tournées vers la simulation des activités humaines supérieures à créé les conditions d'un renouvellement complet des anciens langages de manipulation des symboles du type COMIT, LISP, etc. (list processing languages). En effet, il existe maintenant une nouvelle classe de langages, nés avec les études sur la représentation du savoir, la programmation automatique, la construction d'automates et comportant des caractéristiques complètement nouvelles dans le domaine de la programmation.

On peut les résumer, suivant les subdivisions de Bobrow et Raphael (Bobrow, Raphael 1974), en quatre points.

4.1.1. Nouvelles structures pour l'organisation des données.

L'analyste peut, en effet, définir *a priori*, outre les structures en liste, arbre binaire et chaîne dont il disposait déjà, d'autre structures plus complexes comme, en particulier, des triplés ou des n-tuplés ordonnés, des ensembles non-ordonnés et des affirmations codées selon des symbolismes empruntés à la logique formelle.

4.1.2. Possibilités nouvelles pour le contrôle du déroulement des programmes. Les langages de programmation ordinaires présentent toujours un

contrôle de type hiérarchique, où chaque module logique du programme – lorsqu'il a fini sa tâche – remet le contrôle au module qui l'a appelé. L'innovation fondamentale dans ce domaine, est la possibilité de conserver un module et son contexte en état d'attente. Il n'est plus nécessaire qu'un module aille jusqu'au bout de son déroulement et rende le contrôle au module d'appel, il lui est aussi possible de transmettre le contrôle a quelque autre module en état d'attente qui va reprendre son exécution à partir du point où il avait été suspendu auparavant, sans être obligé, lui non plus, de rendre, à quelque moment, le contrôle à l'un de ses modules d'origine.

Cette innovation dont nous n'avons fait qu'évoquer quelques traits permet en somme le maximum de liberté dans la mise en marche des différents modules en interrompant et en reprenant leur fonctionnement indépendamment de leur organisation hiérarchique de départ.

Nous avons déjà cité une réalisation particulière de cette nouvelle structure de contrôle qui est la possibilité de construire des « demons » c'est-à-dire des modules logiques activés automatiquement lorsque certaines conditions sont réalisées.

4.1.3. « Pattern matching ». Cette possibilité était déjà présente sous une forme assez rigide dans des langages comme COMIT et SNOBOL pour rendre possible l'identification d'une sous-chaîne à l'intérieur d'une chaîne en fonction de son contenu plutôt que de son adresse. Les nouveaux langages permettent entre autre de moduler cette reconnaissance. Un « match », par exemple, peut être retenu comme valable à condition qu'il satisfasse un ensemble de critères spécifié par une variable. Une autre possibilité offerte à l'utilisateur est celle d'obtenir à la demande, aussi bien tous les « matchs » possibles en une fois, qu'une démarche pas à pas où l'exécution de chaque « match » est retardée jusqu'à l'aboutissement de la routine déclenchée par le « match » précédent.

4.1.4. Mécanismes déductifs. Il s'agit de caractéristiques qui permettent dans une certaine mesure l'adaptation automatique – en vue de la solution d'un problème particulier – de la structure algorithmique générale imposée au départ. Le mot *déductif* vient de ce que cette possibilité d'adaptation est étroitement liée aux procédures implémentées dans les démonstrations automatiques de théorems, qui cherchent à déduire les expressions logiques désirées (théorems) à partir de certaines expressions spécifiées au départ (axiomes).

4.2. Les langages qui approchent le mieux ces caractéristiques « idéales » de la programmation A. I. sont le PLANNER – dont seule la version MICRO-PLANNER de Sussman et Winograd est implémentée à ce jour et le CONNIVER – apparemment moins facile à utiliser, car il est plus directif, mais évitant de ce fait à l'usager les risques d'un vagabondage aveugle à travers les embranchements des programmes.

Il ne fait pas de doute que leur utilisation permet d'améliorer considérablement l'efficacité de la programmation A. I.; malheureusement ni PLANNER, ni CONNIVER ne sont implémentés sur la version TSO utilisée au Centre Inter Régional de Calcul Electronique d'Orsay et donc disponibles pour nous – rappelons que RESEDA est construit et sera exploité sur terminal conversationnel 2741 IBM, relié au C.I.R.C.E. –. Cela nous a amené à porter notre choix sur le langage APL/360-TSO.

En effet, des quatre points cités plus haut, deux nous semblent particulièrement indispensables: la possibilité de disposer de nouvelles structures de données et le contrôle souple de déroulement des programmes.

En ce qui concerne le premier point, la structuration en tableaux multidimensionnels qu'APL impose aux données traitées s'est révélée, aux premiers essais, parfaitement adaptée à la représentation des données propres à RESEDA. Ainsi, les parenthétisations emboîtables en triplés contenues dans le plan sont représentées comme des matrices où le nombre de lignes est constant, et égal au nombre total des corrélateurs et où le nombre des colonnes varie, au contraire, en fonction du niveau de parenthétisation de l'épisode enregistré. Dans les structures de parenté, d'autre part, une matrice est réservée à chacune des quatre relations: fraternité, filiation, alliance par fraternité, alliance par filiation. Les éléments sont définis dans chaque matrice en choisissant à l'intersection des deux faisceaux de coordonnées cartésiennes « mâle » et « femelle » certains points significatifs par rapport au mariage que représente le point [0;0] de la matrice filiation. Ces quatre matrices forment ensemble un tableau à trois dimensions où les éléments occupant la même position sur leurs « plans » respectifs – c'est-à-dire ayant leurs deux dernières coordonnées égales – sont dans la même situation par rapport au mariage central (par exemple les « fils » et « fille » de ce mariage, seront les « frère » et « soeur » au plan de la fraternité).

La simulation des nouvelles structures de contrôle de programme, risque de nous poser davantage de problèmes. Néanmoins l'absence en APL d'un véritable concept de « programme » et son remplacement par celui de « fonctions logiques élémentaires » offre sans doute un terrain plus favorable à l'élaboration de structures de programme dépourvues de toute hiérarchie.

LISTE DES OUVRAGES CITES

BOBROW, D. G., RAPHAEL, B. (1974), *New Programming Languages for Artificial Intelligence Research*, « ACM Computing Surveys », VI, 155-174.

BRUCE, B. (1975), *Case Systems for Natural Language*, « Artificial Intelligence », VI, 327-360.

CARBONELL, J. R., COLLINS, A. M. (1974), *Natural Semantics in Artificial Intelligence*, « American Journal of Computational Linguistics », I, microfiche 3.

CECCATO, S. (1962), *NATO Advanced Study Institute on Automatic Translation of Languages - Four Lectures*, Milano, Centro di Cibernetica dell'Università.

CHARNIAK, E. (1972), *Toward a model of children's story comprehension* (AI-TR266 Memo), Cambridge (Mass.), MIT Artificial Intelligence Laboratory.

CHARNIAK, E. (1975), *Organisation and Inference in a Frame-Like System of Common Sense Knowledge*, Castagnola (Lugano), Fondazione Dalle Molle.

CROS, R. C., GARDIN, J. C., LEVY, F. (1968), *L'automatisation des recherches documentaires - Un modèle général, le SYNTOL*, Paris, Gauthier-Villars.

FOSSIER, LUCIE, CREHANGE, MARION (1970), *Un essai de traitement sur ordinateur de documents diplomatiques du Moyen-Age*, « Annales, E.S.C. », XXV, 249-284.

LINDSAY, R. K. (1973), *In defense of Ad Hoc Systems*, in *Computer Models of Thought and Language*, R. C. Schank, K. M. Colby, éds., San Francisco, W. H. Freeman.

MEL'CHUK, I. A. ZHOLKOVSKIJ, A. K. (1970), *Sur la synthèse sémantique*, « T. A. Informations », n. 2, 1-85.

MINSKY, M. (1974), *A Framework for Representing Knowledge* (AI-306 Memo), Cambridge (Mass.), MIT Artificial Intelligence Laboratory.

QUILLIAN, M. R. (1968), *Semantic Memory*, in *Semantic Information Processing*, M. Minsky, éd., Cambridge (Mass.), MIT Press.

SCHANK, R. C. (1973), *Identification of Conceptualizations Underlying Natural Language*, in *Computer Models of Thought and Language*, cit.

SCHANK, R. C., GOLDMAN, N. M., RIEGER, C. J. III, RIESBECK, C. K. (1975), *Inference and Paraphrase by Computer*, « Journal of the Association for Computing Machinery », XXII, 309-328.

SCHWARCZ, R. M., BURGER, J. F., SIMMONS, R. F. (1970), *A Deductive Question Answerer for Natural Language Inference*, « Communications of the ACM », XIII, 167-183.

SCRAGG, G. W. (1975), *Frames, Planes and Nets: A Synthesis*, Castagnola (Lugano), Fondazione Dalle Molle.

WINOGRAD, T. (1972), *Understanding Natural Language*, New York, Academic Press.

WINOGRAD, T. (1973), *A Procedural Model of Language Understanding*, in *Computer Models of Thought and Language*, cit.